Echo-canceling apparatus, an echo-canceling method, a program and a recording medium

BACKGRTOUND OF THE INVENTION

1. Field of the invention

The present invention relates to echo-canceling apparatus comprising a loudspeaker which outputs a received voice from a far-end speaker, a microphone to which the voice of a near-end speaker is input, and a central processing unit (CPU) which controls the whole system, an echo-canceling method for the echo-canceling apparatus as well as a program for the echo-canceling apparatus and a computer-readable recording medium on which the program is recorded.

2.Description of the related art

Voice hands-free apparatus such as a speaker-phone telephone set employs an echo cancellation technique in order to prevent singing and acoustic echo. According to the acoustic echo cancellation technique, from the voice output from a loudspeaker and input as an acoustic echo to a microphone via an acoustic echo path such as a room, the echo replica synthesized in accordance with the echo characteristic is subtracted to substantially cancel the echo.

The related art echo cancellation technique is described below. Fig. 6 is a functional block diagram showing related

art echo-canceling apparatus.

In Fig. 6, a numeral 601 represents a loudspeaker for regenerating a received voice (voice from a far-end speaker) on a speaker phone telephone set, 602 a microphone for picking up the transmitted voice (voice from a near-end speaker), 603 a first echo canceller for canceling the echo propagated over a direct transmission path, 604 a double-talk detector for detecting a double-talk state by using an output signal from the first echo canceller 603, and 605 a second echo canceller for canceling the echo propagated over an indirect transmission path.

The above echo-canceling apparatus may fail to deliver its full performance and become unstable depending on the surrounding noise. As a result, it is difficult to set the learning timing of the first echo canceller, which results in the unstable behavior at the start of conversation. Further, it is difficult to radically suppress singing and automatic recovery is disabled thus releasing an ongoing call.

SUMARRY OF THE INVENTION

In view of the aforementioned problems, the invention aims at providing echo-canceling apparatus which allows conversation immediately following a singing and which delivers a favorable echo cancellation performance from the start of conversation, an echo-canceling method for the echo-

canceling apparatus as well as a program for the echo-canceling apparatus and a computer-readable recording medium on which the program is recorded.

In order to solve the problems, the echo-canceling apparatus of the invention comprises a loudspeaker which outputs a received voice from a far-end speaker, a microphone to which the voice of a near-end speaker is input, and a CPU which controls the whole system, characterized in that the CPU comprises transfer function estimation means which estimates the transfer function of the acoustic echo path between a loudspeaker and a microphone, first filter means which operates using the transfer function estimated by the transfer function estimation means, first subtraction means which subtracts the output signal of the first filter means from the signal from the microphone, second filter means which operates using the transfer function copied from the first filter means in case the estimation accuracy of the transfer function estimation means is high, second subtraction means which subtracts the output signal of the second filter means from the signal from the microphone, singing detection means which detects singing, notch filter means which notches a specific frequency band component in the signal received from a far-end speaker, and switch means which selects between the signal from the far-end speaker processed by the notch filter means and the signal from the far-end speaker not processed by the

notch filter means. This provides echo-canceling apparatus which allows conversation immediately following a singing event and which delivers a favorable echo cancellation performance from the start of conversation.

BRIEF DESCRIPTION OF THE DRAWIINGS

Fig. 1 is a block diagram showing the basic configuration of echo-canceling apparatus according to Embodiment 1 of the invention;

Fig. 2 is a block diagram showing the CPU of echo-canceling apparatus according to Embodiment 1 of the invention;

Fig. 3 is a flowchart showing the operation of the CPU in Fig. 2;

Fig. 4 is a block diagram showing the CPU of echo-canceling apparatus according to Embodiment 2 of the invention;

Fig. 5 is a flowchart showing the operation of the CPU in Fig. 4; and

Fig. 6 is a block diagram showing related art echo-canceling apparatus.

DESCRIPTION OF THE PREFERRED EMBODIMENT

Embodiments of the invention are described below with reference to Figs. 1 through 5.

(Embodiment 1)

Fig. 1 is a block diagram showing the basic configuration of echo-canceling apparatus according to Embodiment 1 of the invention. Fig. 2 is a block diagram showing the CPU of echo-canceling apparatus according to Embodiment 1 of the invention. Fig. 2 shows an echo-canceling method for the echo-canceling apparatus according to Embodiment 1 of the invention. Fig. 3 is a flowchart showing the operation of the CPU in Fig. 2. This feature shows the outline of a program recorded on a ROM.

In Fig. 1, a numeral 101 represents a telephone circuit having an interface to a telephone line, 102 an A/D converter for converting received voice electric signal as an analog electric signal to a digital electric signal, 103 a D/A converter for converting a digital electric signal to an analog electric signal, 104 a loudspeaker for converting an analog electric signal from the D/A converter 103 to a voice, 105 a microphone for converting a voice to an analog electric signal, 106 an A/D converter for converting an analog electric signal from the microphone 105 to a digital electric signal, 107 a D/A converter for converting a digital electric signal to an analog electric signal (transmitted voice electric signal), 108 a CPU for performing digital processing on a digital electric signal from the A/D converter 102 and the A/D converter 106 and outputting the operation result to the D/A

converter 103 and the D/A converter 107, 109 a Read-Only Memory (ROM) where a program to operate the CPU 108 is stored, 110 a Random Access Memory (RAM) used by the CPU 108 as it operates in accordance with the program stored in the ROM 109.

In Fig. 2, a numeral 201 represents singing detection means for detecting singing. The singing detection means 201, detecting a frequency band having a protruding section in the frequency spectrum of a signal from a far-end speaker (hereinafter referred to as a received voice), determines that singing has been made in the frequency band having the protruding section. A numeral 202 represents notch filter means of the band stop type for notching a specific frequency band component, 203 transfer function estimation means for estimating an impulse response of the acoustic echo path between the loudspeaker 104 and the microphone 105 by way of the Steepest Descent Method such as the normalized Least Mean Square (NLMS) method, 204, 205 first and second filter means for performing convolutional operation of the estimated impulse response and the received voice, 206, 207 first and second subtraction means for subtracting the output signals of the first and second filter means from the signal received from the near-end speaker (hereinafter referred to as a transmitted voice), and 208 switch means for selecting whether the received voice will pass through the notch filter means 202 based on the detection result of the singing detection

means 201.

Operation of the CPU 108 thus configured is described below referring to Fig. 3.

In Fig. 3, the transfer function estimation means 203 estimates an impulse response and outputs the estimated response to the first filter means 204. The first filter means 204 performs convolutional operation of the impulse response input from the transfer function estimation means 203 and the received voice, and outputs the operation result to the first subtraction means 206. The first subtraction means 206 subtracts the operation result input from the first filter means 204 from the transmitted voice input from the microphone 105 and outputs the subtraction result to the transfer function estimation means 203 (step 301). The transfer function estimation means 203 monitors the subtraction result input from the first subtraction means 206 (step 302).

In case the estimation accuracy of the transfer function estimation means 203 is low and the subtraction result input from the first subtraction means 206 is unstable, execution returns to step 301.

On the other hand, in case the estimation accuracy of the transfer function estimation means 203 is high and the subtraction result input from the first subtraction means 206 is stable, the second filter means 205 copies and stores a filter coefficient representing an impulse response used by

the first filter means 204 (step 303).

In case the singing detection means 201 has performed singing detection (step 304) and has not detected singing, execution returns to step 301. The second filter means 205 uses the filter coefficient stored in step 303 to perform convolutional operation of the impulse response and the received voice, and outputs the result of convolutional operation to the second subtraction means 207. The second subtraction means 207 subtracts the operation result input from the second filter means 205 from the transmitted voice input from the microphone 105 and outputs the echo-canceled transmitted voice to the D/A converter toward the far-end speaker.

In case the singing detection means 201 has detected singing, the switch means 208 is switched to the notch filter 202 and the received voice is output to the D/A converter 103 at the near-end speaker via the notch filter means 202 (step 305). Copying of the filter coefficient from the first filter means 204 to the second filter means 205 is stopped by the singing detection means 201(step 306). The second filter means 205 continues echo cancellation by using a stored filter coefficient before the singing detection means detected singing. The first filter means 204 initializes the filter coefficient (step 307). In case estimation of an impulse response uses a normalized NLMS, initialization of the filter

coefficient is resetting the filter coefficient to zero (0). The transfer function estimation means 203 resumes leaning from the state where the filter coefficient of the first filter means 204 is reset to 0 and approximates an impulse response in accordance with the subtraction result of the first subtraction means 206 (step 308). When the learning is complete, execution returns to step 301 (step 309).

The notch filter means 202 may be provided as a frequency-variable type and control may be performed so that the notched frequency band will match the frequency band detected by the singing detection means 201 where singing is made.

While estimation of a transfer function uses the Steepest Decent Method (NLMS) method in this embodiment, other methods may be used to estimate a transfer function.

As mentioned hereinabove, this embodiment comprises transfer function estimation means 203 which estimates the transfer function of the acoustic echo path between a loudspeaker 104 and a microphone 105, first filter means 204 which operates using the transfer function estimated by the transfer function estimation means 203, first subtraction means 206 which subtracts the output signal of the first filter means 204 from the signal from the microphone 105, second filter means 205 which operates using the transfer function copied from the first filter means 204 in case the

estimation accuracy of the transfer function estimation means is high, second subtraction means 207 which subtracts the output signal of the second filter means 205 from the signal from the microphone 105, singing detection means 201 which detects singing, notch filter means 202 which notches a specific frequency band component in the signal received from a far-end speaker, and switch means 208 which selects between the signal from the far-end speaker processed by the notch filter means 202 and the signal from the far-end speaker not processed by the notch filter means 202. A singing frequency is filtered out by the notch filter means 202 on detection of singing and the transfer function stored before detection of singing is used to perform echo cancellation. This allows conversation immediately following a singing event. On detection of singing, the transfer function of the first filter means 204 is initialized. The signal from the far-end speaker where a singing frequency component has been removed by the notch filter means 202 is used to learn the transfer function. Once learning of the transfer function is complete, the transfer function is copied from the first filter means 204 to the second filter 205. This delivers a favorable echo cancellation performance from the start of conversation.

Running a program to execute the steps of the echo-canceling method shown in Fig. 3 on a computer allows execution of the echo-canceling method of this embodiment in

an arbitrary place at an arbitrary time. By reading on a computer a recording medium where the program is recorded, it is possible to execute the program in an arbitrary place at an arbitrary time.

(Embodiment 2)

Fig. 4 is a functional block diagram showing the CPU of echo-canceling apparatus according to Embodiment 2. Fig. 5 is a flowchart showing the operation of the CPU in Fig. 4. The basic configuration of the echo-canceling apparatus according to this embodiment is the same as that shown in Fig. 1. This feature shows the outline of a program recorded on a ROM.

In Fig. 4, a numeral 401 represents speaker detection means which detects the speech of a far-end speaker, speech of a near-end speaker and a double-talk (simultaneous speech of the far-end speaker and the near-end speaker), 402 transfer function estimation means which estimates the transfer function of the acoustic echo path between a loudspeaker 104 and a microphone 105 by way of the Steepest Descent Method such as the normalized Least Mean Square (NLMS) method, 403 direct echo filter means which performs convolutional operation of a transfer function corresponding to a direct echo component and a received voice, 404 indirect echo filter means which performs convolutional operation of a transfer function corresponding to an indirect echo component and the received voice, and 405 subtraction means.

The direct echo component refers to a voice emitted from the loudspeaker 104 and directly input to the microphone 105. The indirect echo component refers to a voice emitted from the loudspeaker 104, reflected against objects such as a wall, a floor and a ceiling in an acoustic echo path, and input to the microphone 105.

General operation of the echo-canceling apparatus thus configured is described below referring to Fig. 5.

In Fig. 5, when echo cancellation is started (step 501), the speaker detection means 401 determines whether the talking state is speech of the far-end speaker, speech of the near-end speaker or double talk (step 502). In case the talking state is speech of the far-end speaker, the transfer function estimation means 402 uses an algorithm such as NLMS to estimate a direct echo component transfer function (step 503) and an indirect echo component transfer function (step 504). The direct echo filter means 403 performs convolutional operation of the result of the estimation of direct echo component transfer function (step 503) and a received voice (step 505) while the indirect echo filter means 404 performs convolutional operation of the result of estimation of indirect echo component transfer function (step 504) and the received voice (step 506). The result of convolutional operation is subtracted from the transmitted voice from the microphone 105 on the subtraction means 405 to remove the

direct echo component and the indirect echo component (step 507).

This provides echo cancellation which allows high-speed and high-accuracy estimation of a transfer function.

As mentioned hereinabove, according to this embodiment, the direct echo filter means 403 performs convolutional operation of the result of the estimation of direct echo component transfer function (step 503) and a received voice while the indirect echo filter means 404 performs convolutional operation of the result of estimation of indirect echo component transfer function (step 504) and the received voice. The result of convolutional operation is subtracted from the transmitted voice from the microphone 105 on the subtraction means 405 to remove the direct echo component and the indirect echo component. This maintains high the double talk determination accuracy even in case the volume of the voice from the loudspeaker is increased. Double talk detection accuracy is maintained high even in case the voice power ratio of the received voice and the transmitted voice is the same.

CROSS REFERENCE TO RELATED APPLICATION

This application is based upon and claims the benefit of priority of Japanese Patent Application No2003-066481 filed on 03/12/03, the contents of which are incorporated herein by

reference in its entirety.